Task 1
Create a database named 'custom'.
Create a table named temperature_data inside custom having below fields:
1. date (mm-dd-yyyy) format
2. zip code
3. temperature
The table will be loaded from comma-delimited file.
Load the dataset.txt (which is ',' delimited) in the table.

COMMAND:-

(i)create database custom;

(ii) create table temperature_data
(
dates STRING,
zip_code INT,
temperature SMALLINT
)

row format delimited fields terminated by ',';
[note:-Later I use type casting over STRING data types of dates]

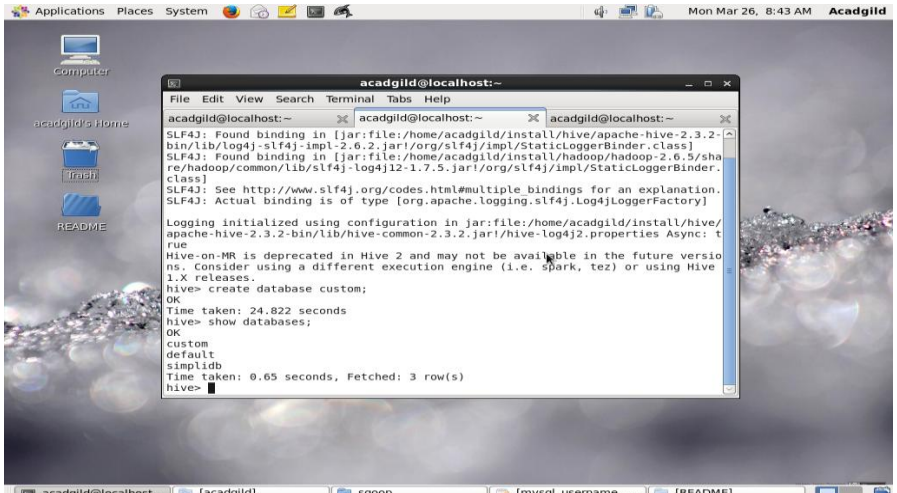
(iii) LOAD DATA LOCAL INPATH
'/home/acadgild/hive_folder/dataset_Session.txt' into table
temperature_data;

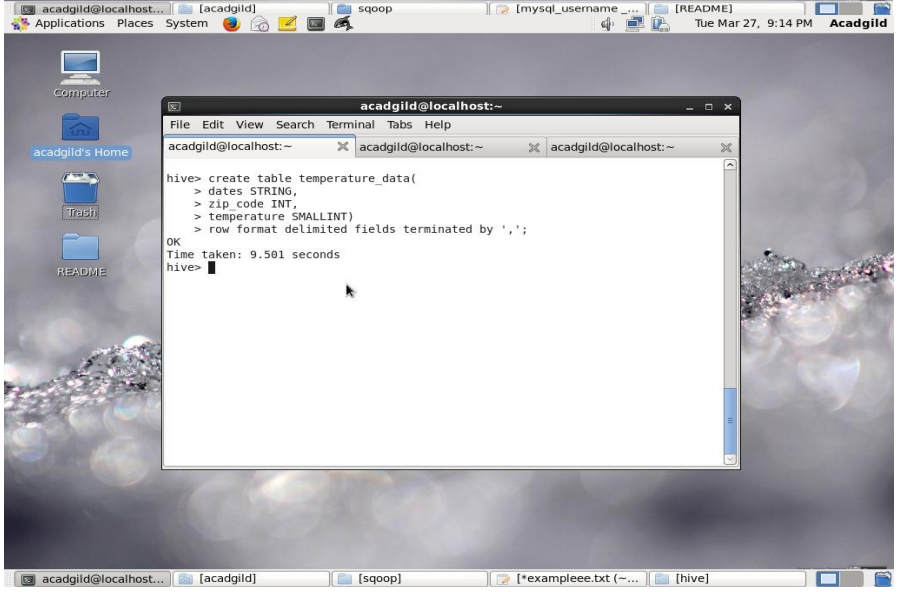
(iv) select * from temperature_data;

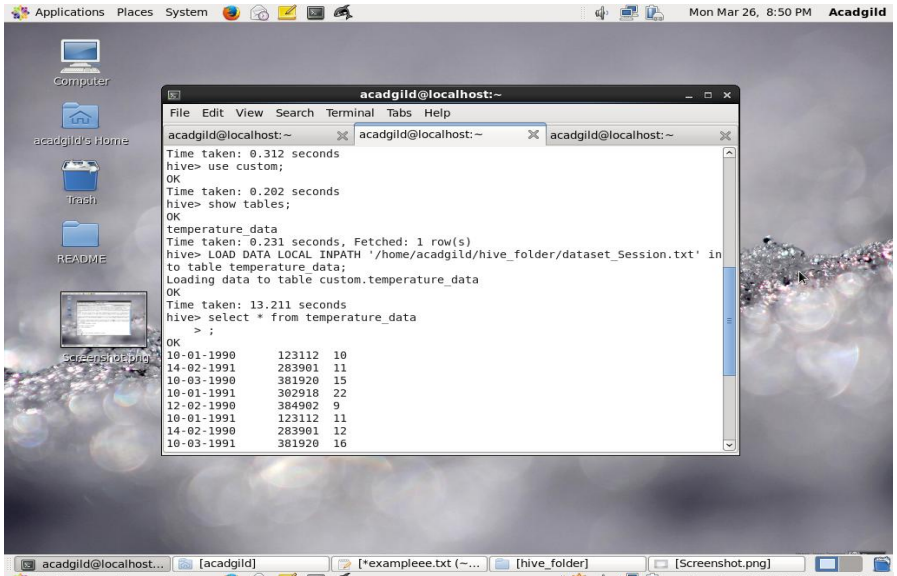
[note:- here I cast the string data types of dates to date]
(v)select cast(to_date(from_unixtime(unix_timestamp(dates,
'dd-MM-yyyy'))) as date),zip_code,temperature from
temperature_data;

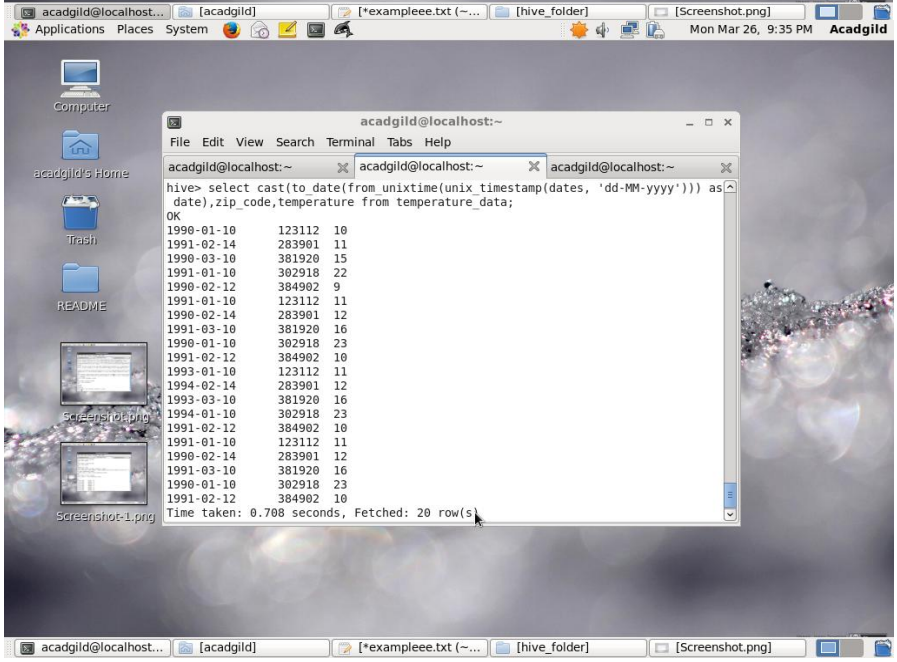
EXPLANATION:-
(i) create the custom database here.
(ii)create the temperature_data here.
(iii)load the content of the dataset_Session.txt the temperature_data table.
(iv) check the temperature_data tables contents.
(v)after type casting check the contents of temperature_data again.

OUTPUT:-


Applications Places System
Mon Mar 26, 8:43 AM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
acadgild@localhost:~
acadgild@localhost:~
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> create database custom;
OK
Time taken: 24.822 seconds
hive> show databases;
OK
custom
default
simplidb
Time taken: 0.65 seconds, Fetched: 3 row(s)
hive>
acadgild@localhost...
[acadgild]
sqoop
[mysql_username _...
[README]


Applications Places System
Tue Mar 27, 9:14 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
acadgild@localhost:~
acadgild@localhost:~
hive> create table temperature_data(
> dates STRING,
> zip_code INT,
> temperature SMALLINT)
> row format delimited fields terminated by ',';
OK
Time taken: 9.501 seconds
hive>
acadgild@localhost...
[acadgild]
[sqoop]
[*exampleee.txt (~...
[hive]





Task 2

- Fetch date and temperature from temperature_data where zip code is greater than 300000 and less than 399999.
- Calculate maximum temperature corresponding to every year from temperature_data table.
- Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.
- Create a view on the top of last query, name it temperature_data_vw.
- Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.

COMMAND:-

(i) select dates,temperature,zip_code from temperature_data where zip_code>300000 and zip_code<399999;

(ii) select YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)),MAX(temperature) from temperature_data group by YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date));

(iii) select YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)),MAX(temperature) from temperature_data group by YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)) having COUNT(YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)))>=2;

(iv) CREATE VIEW temperature_data_vw As select YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)),MAX(temperature) from temperature_data group by YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)) having COUNT(YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)))>=2;

(v) hive -S -e "USE custom; select * from temperature_data_vw" | sed 's/[\t]/|/g' > /home/acadgild/test.txt

EXPLANATION:-

(i) Fetch date, temperature and zip_codes from temperature_data where zip code is greater than 300000 and less than 399999.

(ii) Calculate maximum temperature corresponding to every year from temperature_data table.

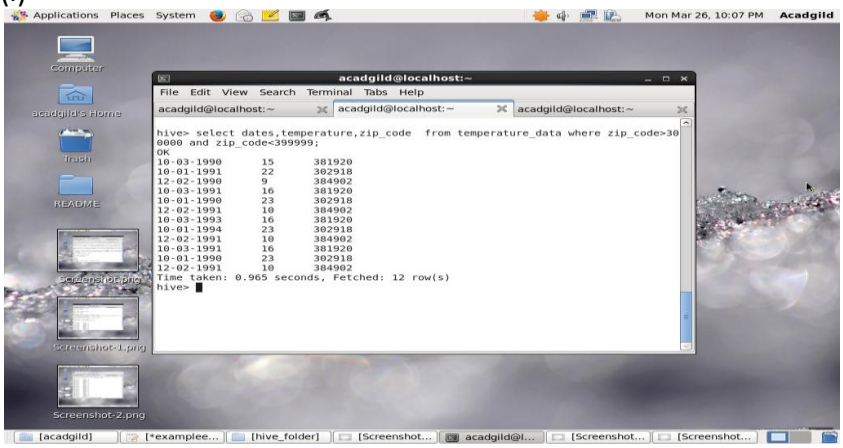
(iii) Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.

(iv) Create a view on the top of last query, name it temperature_data_vw.

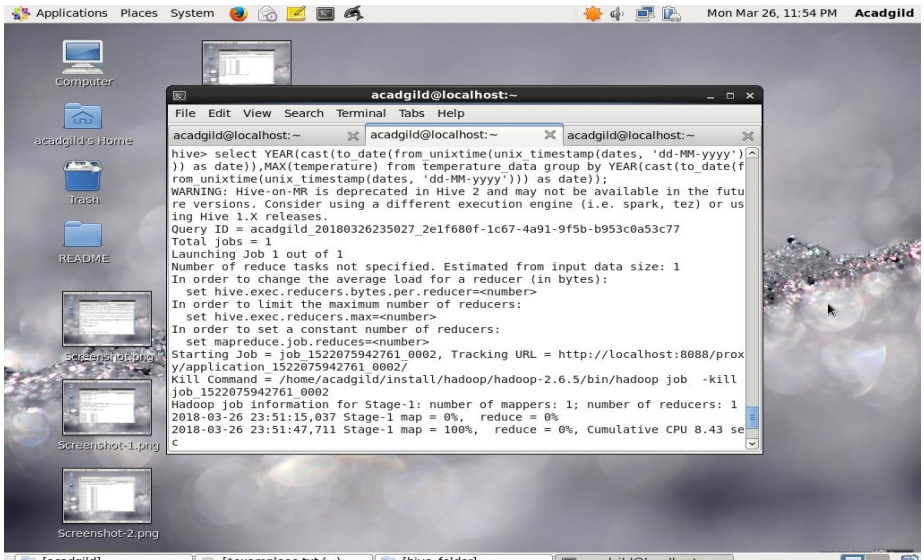
(v) Export contents from temperature_data_vw to a file test.txt in local file system, such that each file is '|' delimited.

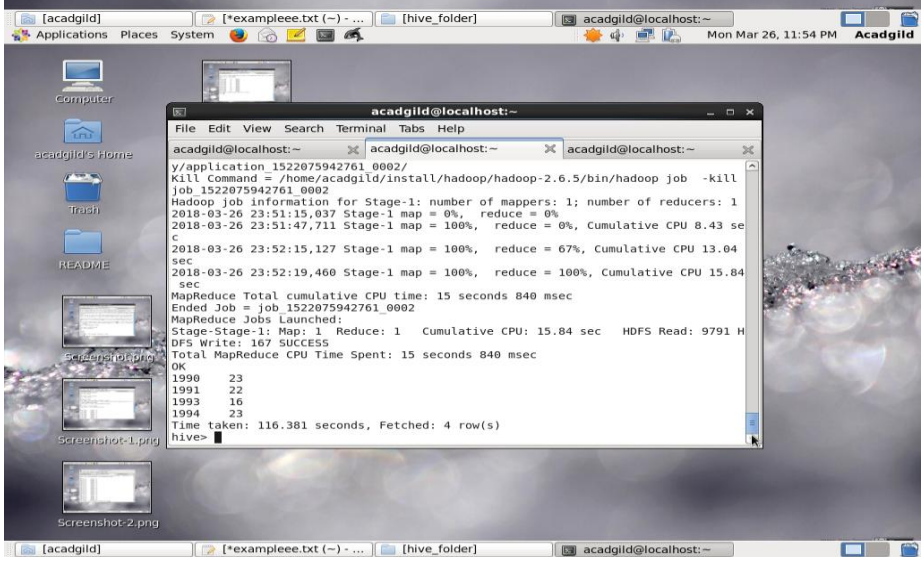
OUTPUT:-

(i)

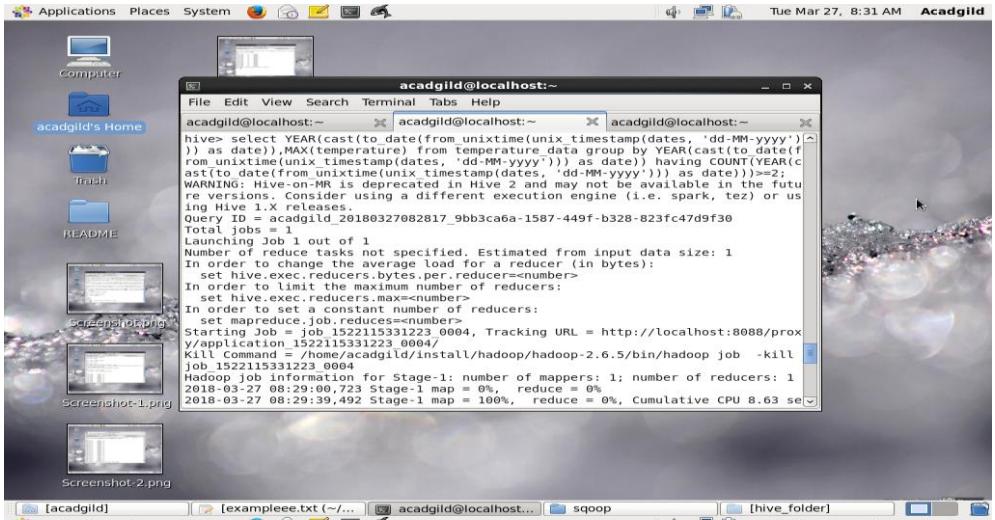

Applications Places System
Mon Mar 26, 10:07 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
hive> select dates,temperature,zip_code from temperature_data where zip_code>30
0000 and zip_code<399999;
OK
10-03-1990 15 381920
10-01-1991 22 302918
12-02-1990 9 384902
10-03-1991 16 381920
10-01-1990 23 302918
12-02-1991 10 384902
10-03-1993 16 381920
10-01-1994 23 302918
12-02-1991 10 384902
10-03-1991 16 381920
10-01-1990 23 302918
12-02-1991 10 384902
Time taken: 0.965 seconds, Fetched: 12 row(s)
hive>

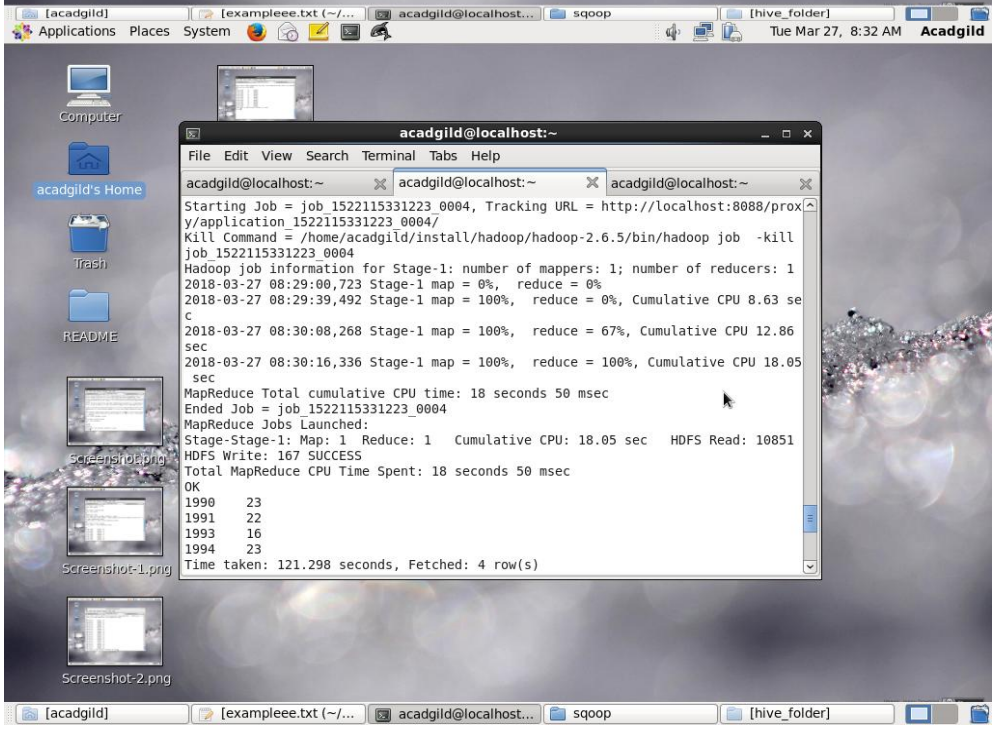
(ii)


Applications Places System
Mon Mar 26, 11:54 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
hive> select YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy')
)) as date)),MAX(temperature) from temperature_data group by YEAR(cast(to_date(f
rom_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date));
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the futu
re versions. Consider using a different execution engine (i.e. spark, tez) or us
ing Hive 1.X releases.
Query ID = acadgild_20180326235027_2e1f680f-1c67-4a91-9f5b-b953c0a53c77
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapreduce.job.reduces=<number>
Starting Job = job_1522075942761_0002, Tracking URL = http://localhost:8088/prox
y/application_1522075942761_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill
job_1522075942761_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-03-26 23:51:15,037 Stage-1 map = 0%, reduce = 0%
2018-03-26 23:51:47,711 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.43 se
c

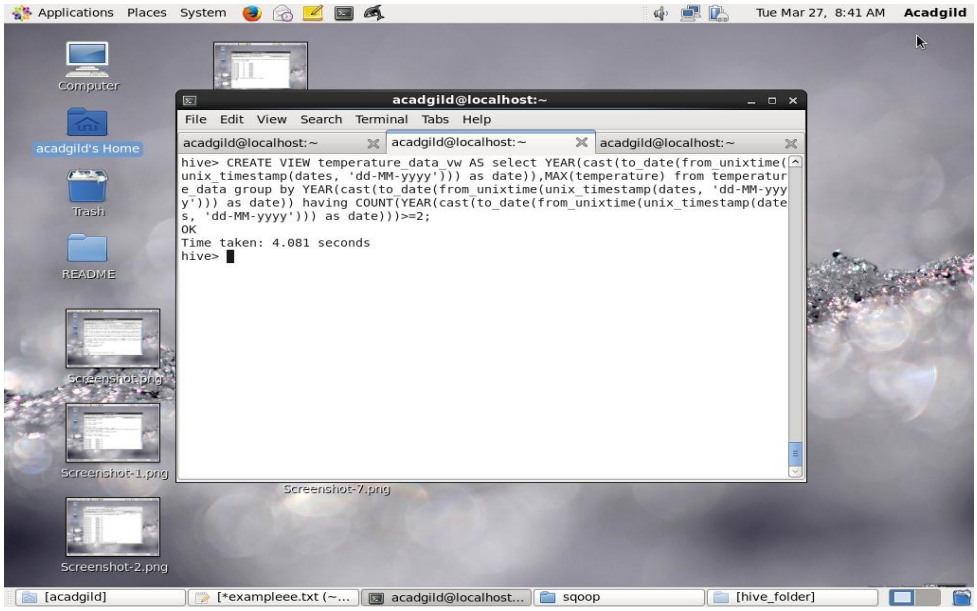

Applications Places System
Mon Mar 26, 11:54 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
y/application_1522075942761_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill
job_1522075942761_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-03-26 23:51:15,037 Stage-1 map = 0%, reduce = 0%
2018-03-26 23:51:47,711 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.43 se
c
2018-03-26 23:52:15,127 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 13.04
sec
2018-03-26 23:52:19,460 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 15.84
sec
MapReduce Total cumulative CPU time: 15 seconds 840 msec
Ended Job = job_1522075942761_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 15.84 sec HDFS Read: 9791 H
DFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 840 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 116.381 seconds, Fetched: 4 row(s)
hive>

(iii)

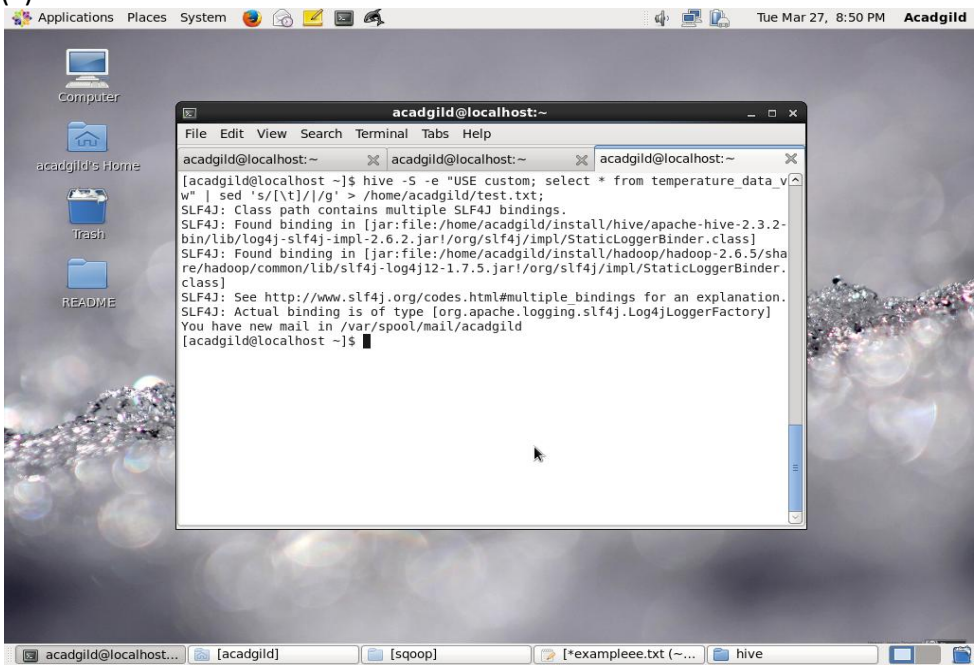

Applications Places System
Tue Mar 27, 8:31 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
hive> select YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)),MAX(temperature) from temperature_data group by YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)) having COUNT(YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyyy'))) as date)))>=2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180327082817_9bb3ca6a-1587-449f-b328-823fc47d9f30
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapreduce.job.reduces=<number>
Starting Job = job_1522115331223_0004, Tracking URL = http://localhost:8088/proxy/application_1522115331223_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1522115331223_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-03-27 08:29:00,723 Stage-1 map = 0%, reduce = 0%
2018-03-27 08:29:39,492 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.63 se

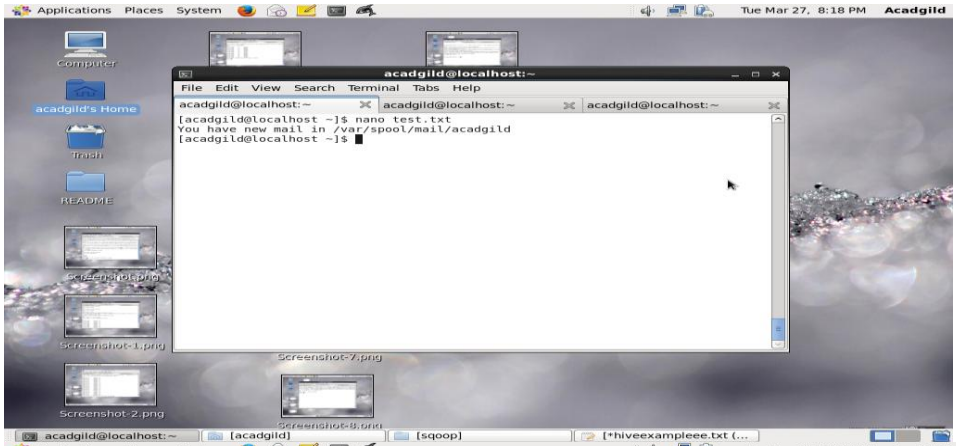

Applications Places System
Tue Mar 27, 8:32 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
Starting Job = job_1522115331223_0004, Tracking URL = http://localhost:8088/proxy/application_1522115331223_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1522115331223_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-03-27 08:29:00,723 Stage-1 map = 0%, reduce = 0%
2018-03-27 08:29:39,492 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.63 sec
2018-03-27 08:30:08,268 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.86 sec
2018-03-27 08:30:16,336 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 18.05 sec
MapReduce Total cumulative CPU time: 18 seconds 50 msec
Ended Job = job_1522115331223_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 18.05 sec HDFS Read: 10851 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 50 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 121.298 seconds, Fetched: 4 row(s)

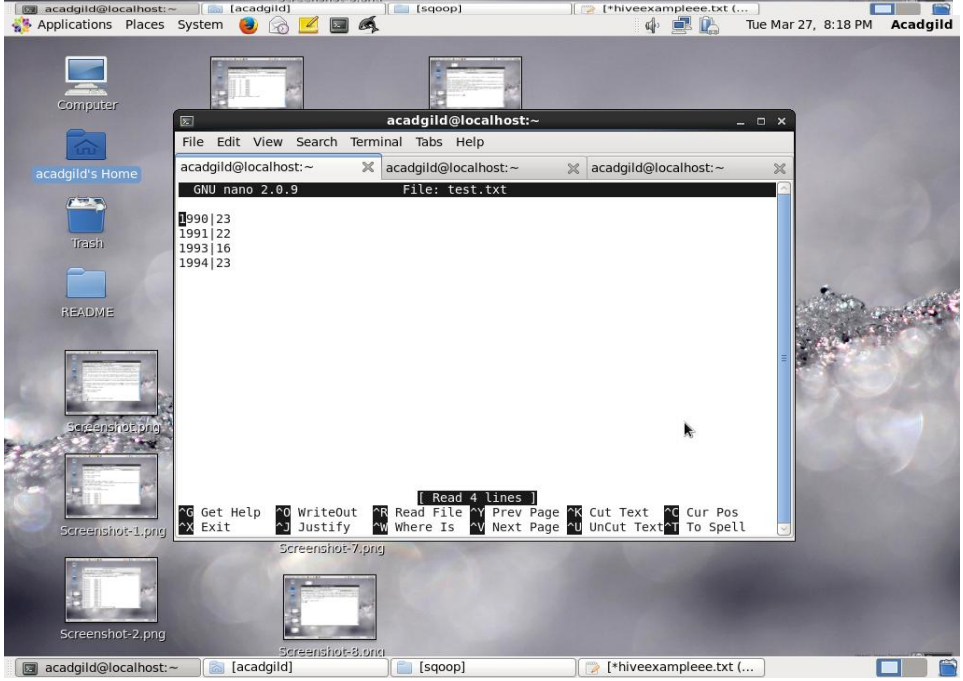
(iv)


Applications Places System
Tue Mar 27, 8:41 AM Acadgild
Computer
acadgild's Home
Trash
README
Screenshot.png
Screenshot-1.png
Screenshot-2.png
Screenshot-7.png
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
hive> CREATE VIEW temperature_data_vw AS select YEAR(cast(to_date(from_unixtime(
unix_timestamp(dates, 'dd-MM-yyyy'))) as date)),MAX(temperature) from temperatur
e_data group by YEAR(cast(to_date(from_unixtime(unix_timestamp(dates, 'dd-MM-yyy
y'))) as date)) having COUNT(YEAR(cast(to_date(from_unixtime(unix_timestamp(date
s, 'dd-MM-yyyy'))) as date)))>=2;
OK
Time taken: 4.081 seconds
hive>
[acadgild]
[*exampleee.txt (~...
acadgild@localhost...
sqoop
[hive_folder]

(v)


Applications Places System
Tue Mar 27, 8:50 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
[acadgild@localhost ~]$ hive -S -e "USE custom; select * from temperature_data_v
w" | sed 's/[\t]/|/g' > /home/acadgild/test.txt;
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-
bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/sha
re/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.
class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
acadgild@localhost...
[acadgild]
[sqoop]
[*exampleee.txt (~...
hive


Applications Places System
Tue Mar 27, 8:18 PM Acadgild
Computer
acadgild's Home
Trash
README
Screenshot.png
Screenshot-1.png
Screenshot-2.png
Screenshot-7.png
Screenshot-8.png
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
acadgild@localhost:~
acadgild@localhost:~
[acadgild@localhost ~]$ nano test.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
acadgild@localhost:~
[acadgild]
[sqoop]
[*hiveexampleee.txt (...


Applications Places System
Tue Mar 27, 8:18 PM Acadgild
Computer
acadgild's Home
Trash
README
Screenshot.png
Screenshot-1.png
Screenshot-2.png
Screenshot-7.png
Screenshot-8.png
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
acadgild@localhost:~
acadgild@localhost:~
GNU nano 2.0.9 File: test.txt
1990|23
1991|22
1993|16
1994|23
[Read 4 lines]
^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text ^C Cur Pos
^X Exit ^J Justify ^W Where Is ^V Next Page ^U UnCut Text ^T To Spell
acadgild@localhost:~
[acadgild]
[sqoop]
[*hiveexampleee.txt (...